where the second equation is obtained because RNN is an accumulative process, and the function $\mathcal{G}$ is determined by the parameters in RNN, learned from the training data. The function $\mathcal{G}$ has high capacities to approximate the functions in the formula of BM25 since there are many parameters. Therefore, a simplified version of DeepRank can well approximate the BM25 model.

In addition, DeepRank has closer relationships with MatchPyramid and Match-SRNN. If we set the window size of query-centric context to be $k = |\mathbf{d}|$ and the weights of query term w_u to be $1/f(w_u, \mathbf{d})$, DeepRank reduces to MatchPyramid or Match-SRNN, by using CNN or 2D-GRU as the measure network, respectively.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we propose a new deep learning architecture, namely DeepRank. Firstly, a detection strategy is designed to extract query-centric contexts. A measure network is then applied to determine the local relevance between query and each query-centric context, by using CNN or 2D-GRU. Finally, an aggregation network is used to produce the global relevance score, via RNN and a term gating network. DeepRank not only well simulates the relevance generation process in human judgement, but also captures important IR characteristics, i.e. exact/semantic matching signals, proximity heuristics, query term importance, and diverse relevance requirement. We conduct experiments on both benchmark LETOR4.0 data and a large clickthrough data. The results show that DeepRank significantly outperform learning to rank methods and existing deep IR models, when most existing deep IR models perform much worse than learning to rank methods. To the best of our knowledge, DeepRank is the first deep IR model to outperform existing learning to rank models. We also give a detailed analysis on DeepRank to show insights on parameter settings for implementation.

For future work, we plan to investigate the differences between the automatically learned representations of DeepRank and effective features used in learning to rank, which may introduce some insights for architecture design of more powerful deep IR models.

7 ACKNOWLEDGMENTS

This work was funded by the 973 Program of China under Grant No. 2014CB340401, the National Natural Science Foundation of China (NSFC) under Grants No. 61232010, 61433014, 61425016, 61472401, and 61203298, and the Youth Innovation Promotion Association CAS under Grants No. 20144310 and 2016102. The authors would like to thank Chengxiang Zhai (UIUC) and Yixing Fan (ICT, CAS) for their valuable suggestions on this work, and Weipeng Chen (Sogou Inc.) for providing helps on the data processing of ChineseClick.

REFERENCES

[1] Christopher JC Burges. 2010. From ranknet to lambdarank to lambdamart: An overview. *Learning* 11 (2010), 23–581.
[2] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. 2007. Learning to rank: from pairwise approach to listwise approach. In *ICML*. ACM, 129–136.
[3] Olivier Chapelle and Yi Chang. 2011. Yahoo! learning to rank challenge overview. In *Proceedings of the Learning to Rank Challenge*. 1–24.
[4] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *EMNLP*. 1724–1734.
[5] Carsten Eickhoff, Sebastian Dungs, and Vu Tran. 2015. An eye-tracking study of query reformulation. In *SIGIR*. ACM, 13–22.
[6] Hui Fang, Tao Tao, and ChengXiang Zhai. 2004. A formal study of information retrieval heuristics. In *SIGIR*. ACM, 49–56.
[7] Yoav Freund, Raj Iyer, Robert E Schapire, and Yoram Singer. 2003. An efficient boosting algorithm for combining preferences. *JMLR* 4, Nov (2003), 933–969.
[8] Fredric C Gey. 1994. Inferring probability of relevance using the method of logistic regression. In *SIGIR*. Springer, 222–231.
[9] Rich Caruana Steve Lawrence Lee Giles. 2001. Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping. In *NIPS*, Vol. 13. MIT Press, 402.
[10] Alan Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. 2013. Speech recognition with deep recurrent neural networks. In *ICASSP*. IEEE, 6645–6649.
[11] Jiafeng Guo, Yixing Fan, Qingyao Ai, and W Bruce Croft. 2016. A deep relevance matching model for ad-hoc retrieval. In *CIKM*. ACM, 55–64.
[12] Baotian Hu, Zhengdong Lu, Hang Li, and Qingcai Chen. 2014. Convolutional neural network architectures for matching natural language sentences. In *NIPS*. 2042–2050.
[13] Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. In *CIKM*. ACM, 2333–2338.
[14] Thorsten Joachims. 2002. Optimizing search engines using clickthrough data. In *SIGKDD*. ACM, 133–142.
[15] Thorsten Joachims. 2006. Training linear SVMs in linear time. In *SIGIR*. ACM, 217–226.
[16] Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
[17] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.
[18] Tie-Yan Liu. 2009. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval* 3, 3 (2009), 225–331.
[19] Yuanhua Lv and ChengXiang Zhai. 2009. Positional language models for information retrieval. In *SIGIR*. ACM, 299–306.
[20] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*. 3111–3119.
[21] Shuzi Niu, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. 2012. Top-k learning to rank: labeling, ranking and evaluation. In *SIGIR*. ACM, 751–760.
[22] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web.* Technical Report. Stanford InfoLab.
[23] Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, and Xueqi Cheng. 2016. A study of matchpyramid models on ad-hoc retrieval. In *Neu-IR fi16 SIGIR Workshop on Neural Information Retrieval*.
[24] Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, Shengxian Wan, and Xueqi Cheng. 2016. Text matching as image recognition. In *AAAI*. AAAI Press, 2793–2799.
[25] Tao Qin, Tie-Yan Liu, Jun Xu, and Hang Li. 2010. LETOR: A benchmark collection for research on learning to rank for information retrieval. *Information Retrieval* 13, 4 (2010), 346–374.
[26] Stephen Robertson. 2000. Evaluation in information retrieval. In *Lectures on information retrieval*. Springer, 81–92.
[27] Stephen E Robertson and Steve Walker. 1994. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *SIGIR*. Springer-Verlag New York, Inc., 232–241.
[28] Aliaksei Severyn and Alessandro Moschitti. 2015. Learning to rank short text pairs with convolutional deep neural networks. In *Proceedings of SIGIR*. ACM, 373–382.
[29] Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng, and Grégoire Mesnil. 2014. Learning semantic representations using convolutional neural networks for web search. In *WWW*. WWW, 373–374.
[30] Mark D Smucker, James Allan, and Ben Carterette. 2007. A comparison of statistical significance tests for information retrieval evaluation. In *CIKM*. ACM, 623–632.
[31] Tao Tao and ChengXiang Zhai. 2007. An exploration of proximity measures in information retrieval. In *SIGIR*. ACM, 295–302.
[32] Shengxian Wan, Yanyan Lan, Jiafeng Guo, Jun Xu, Liang Pang, and Xueqi Cheng. 2016. Match-SRNN: Modeling the Recursive Matching Structure with Spatial RNN. In *IJCAI*. 2922–2928.
[33] Ho Chung Wu, Robert WP Luk, Kam-Fai Wong, and KL Kwok. 2007. A retrospective study of a hybrid document-context based retrieval model. *Information processing & management* 43, 5 (2007), 1308–1331.
[34] Jun Xu and Hang Li. 2007. Adarank: a boosting algorithm for information retrieval. In *SIGIR*. ACM, 391–398.